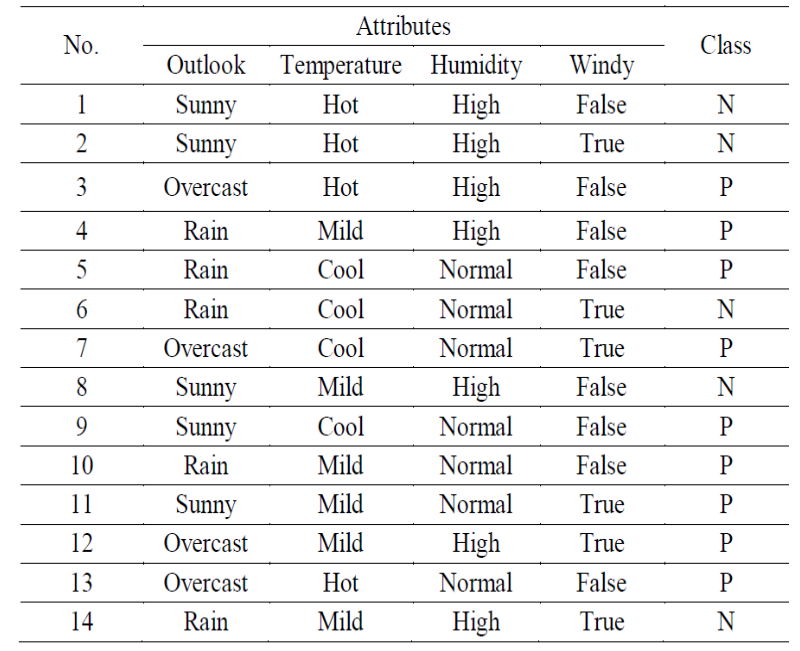
Subject Name: Data Mining and Business Intelligence

Subject Code: 2170715

IMPORTANT QUESTIONS

1	Explain various features of Data Warehouse? Compare data mart and data warehouse.
2	Explain three tier data warehouse architecture.
3	Differentiate between OLAP and OLTP systems
4	What is cuboid? Explain various OLAP operations on data cube with suitable example.
5	Explain Star, Snowflake, Fact Constellation Schema for Multidimensional Database.
6	List and describe major issues in data mining.
7	Define KDD. Explain KDD process in detail.
8	What is the need of data pre-processing? List and describe the methods for handling the missing values in data cleaning.
9	What is noise? Explain data smoothing methods as noise removal technique to divide given data into bins of size 3 by bin partition (equal frequency), by bin means, by bin medians and by bin boundaries. Consider the data: 10, 2, 19, 18, 20, 18, 25, 28, 22
10	Explain the following data normalization techniques. With example. **(1)** Min-max normalization **(2)** Decimal scaling. **(3)** z-score normalization
11	Explain Mean, Median, Mode, Variance, Standard Deviation & five number summary with suitable database example.
12	What is Concept Hierarchy? List and explain types of Concept Hierarchy.
13	Explain with an example attribute removal and attribute generalization.
14	What is Market Basket Analysis? Explain Association Rules with Confidence & Support.
15	What do you mean by frequent item set mining for market basket analysis? Explain apriori algorithm for the same with suitable example.
16	What are the limitations of the Apriori approach for mining? Briefly describe the techniques to improve the efficiency of Apriori algorithm.
17	State the Apriori Property. Generate large itemsets and association rules using Apriori algorithm on the following data set with minimum support value and minimum confidence value set as 50% and 75% respectively. ***TID*** ***Items Purchased*** T101 Cheese, Milk, Cookies T102 Butter, Milk, Bread T103 Cheese, Butter, Milk, Bread T104 Butter, Bread
18	What is supervised learning? Using the given table, show how the ROOT splitting attribute is selected using InfoGain measure in the overall process of decision tree induction.



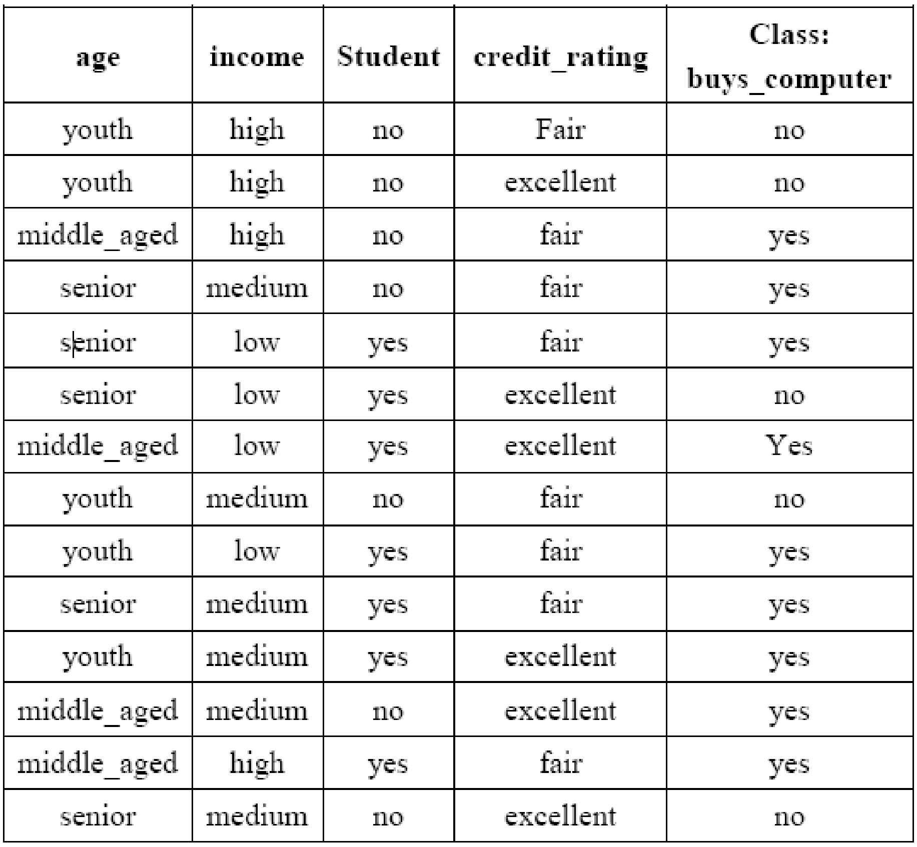
No.	Attributes				Class
	Outlook	Temperature	Humidity	Windy	
1	Sunny	Hot	High	False	N
2	Sunny	Hot	High	True	N
3	Overcast	Hot	High	False	P
4	Rain	Mild	High	False	P
5	Rain	Cool	Normal	False	P
6	Rain	Cool	Normal	True	N
7	Overcast	Cool	Normal	True	P
8	Sunny	Mild	High	False	N
9	Sunny	Cool	Normal	False	P
10	Rain	Mild	Normal	False	P
11	Sunny	Mild	Normal	True	P
12	Overcast	Mild	High	True	P
13	Overcast	Hot	Normal	False	P
14	Rain	Mild	High	True	N

19 What is decision tree induction? Write Basic algorithm for inducing a decision tree from training tuples.

20 Explain Baye's theorem and Naive Bayesian classification.

21 What are neural networks? Describe the various factors which make them useful for classification and prediction in data mining. Explain how the topology of neural network is designed

22 Using Naive Bayesian classification method, predict class label of X = (age = youth, income = medium, student = yes, credit_rating = fair) using following training dataset.



age	income	Student	credit_rating	Class: buys_computer
youth	high	no	Fair	no
youth	high	no	excellent	no
middle_aged	high	no	fair	yes
senior	medium	no	fair	yes
senior	low	yes	fair	yes
senior	low	yes	excellent	no
middle_aged	low	yes	excellent	Yes
youth	medium	no	fair	no
youth	low	yes	fair	yes
senior	medium	yes	fair	yes
youth	medium	yes	excellent	yes
middle_aged	medium	no	excellent	yes
middle_aged	high	yes	fair	yes
senior	medium	no	excellent	no

23	Explain Linear Regression and Non-linear Regression techniques of prediction.
24	How data Mining is useful for Business Intelligence applications viz.Balanced Scorecard, Fraud Detection, Clickstream Mining, Market Segmentation, Retail industry, Telecommunications industry, Banking & Finance and CRM
25	What is Big Data? What is big data analytic? Explain the big data- distributed file system.
26	What is Cluster Analysis? List and explain requirements of clustering in data mining.
27	What is an 'outlier'? How do outliers impact the results of mining? Explain any one method to detect outliers.
28	Explain different types of web mining with suitable example.
29	Explain Text mining using example.
30	Explain Hadoop architecture using figure. Discuss the main features of Hadoop distributed file system.